Journal of Official Statistics
Vol. 4, No. 1, 1988, pp. 53–65

Miscellanea

Under the heading Miscellanea, essays will be published dealing with topics considered to be of general interest to the readers. All contributions will be refereed for their compatibility with this criterion.

Data, Statistics, Information:

Some Issues of the Canadian Social Statistics Scene[1]

Ivan P. Fellegi[2]

1. Introduction

Any discussion on public policy needs insight and information. Too often, we offer statistics as a substitute.

It is the thesis of this paper that there is a vital difference between information and statistics. And the key is analysis: it is analysis that transforms statistics into information; it is the exploitation of the underlying information that renders statistics useful; and it is the recognized usefulness of statistics which is the foundation for the survival of national statistical systems.

Section 2 of this paper lays out some basic concepts of the characteristics of data, statistics, and information. However, Section 2 can be omitted by those who wish to focus on the applied part of the paper. Section 3 is a case study which reviews the current state of social statistics in Canada. Section 4 outlines some recent initiatives by Statistics Canada to promote more effective analysis of statistics inside and outside the agency to improve the relevance of our output. Some concluding remarks end the paper.

[1] This is a substantially revised version of a paper previously published in *Survey Methodology*, 1979, Vol. 5, No. 2, pp. 130–161.

[2] Chief Statistician, Statistics Canada, Ottawa, K1A OTA 6 Canada.

2. Some Basic Concepts

2.1. *Datum*

A datum is a quantity (e.g., the dollar value of sales) or a code (e.g., the numerical code assigned to identify an industry, race, sex, occupation, etc.). However, a somewhat deeper analysis reveals that semantically the definition of datum must include at least three components: a quantity or code value (which can be a vector); a precise definition of what has been quantified (e.g., the particular definition of income used in a survey of consumer finances, or the definition of industry codes in the Canadian Labour Force Survey); and a reference entity, i.e., a precise identification of the object or set of objects to which the quantity or code refers. If any of these three components is time-dependent, the reference time should also be part of the definition of the datum. Finally, the description of the measurement process yielding the quantity or code number should, whenever relevant, be part of the datum definition.

2.2. *Microdata*

A micro datum is a datum referring to a single object. The reference entity in this case is therefore a single person, business,

institution, a parcel of real estate, etc. The notion of an "object" is usually clear in any given context and will not be defined here rigorously.

2.3. *Statistic*

A statistic is a datum resulting from a specified mathematical operation carried out on microdata and refers to an unambiguously defined set of objects (persons, businesses, events, phenomena, etc.). The set over which the mathematical operation is carried out may or may not coincide with the reference set for the statistic. The reference set is a group definition (universe definition using sample survey terminology) independent of the particular operation (e.g., sampling, interviewing) which yielded the microdata underlying the statistic. The transition from the set from which the microdata are drawn and the reference set to which the statistic refers is embodied in the definition of the mathematical operation yielding the statistic. This latter definition, together with the definitions of the variables associated with the microdata comprise the definition component of the statistics. The reference time and the description of the measurement process usually carry over from the underlying microdata.

2.4. *Interpretable message*

A datum or statistic which has been communicated by a person or institution is defined as an interpretable message. For example, by virtue of the dissemination of statistics on the unemployed in Nova Scotia, by age and sex, these statistics become interpretable messages. To cite a more special example, data sent into outer space are interpretable messages, but we do not know whether they are received and interpreted by anyone.

2.5. *Information*

To inform is a process. Information, for the purposes of this paper, is defined as the result of a process of conveying an interpretable message in such a manner that the receiver of the message acquires knowledge, i.e., becomes better informed. Hence information involves interpretation.

An interpretable message has the potential of informing, but it becomes information only if it is received by an intelligent receiver who interprets it; i.e., does not screen it out but stores it in his/her mind for some expected use. An intelligent receiver is a person with the knowledge needed to "decode" a statistic into the three components of data identified earlier. He/she also has the ability to relate the decoded statistic to other aspects of his/her knowledge or experience.

Whether or not an interpretable message, when received, is screened out or stored can be influenced by the sender, e.g., through the medium and presentation used, or through repetition. More importantly, the sender can induce storage and interpretation of the message by calling attention to the fact that the message has some intrinsic interest or utility for the receiver.

2.6. *Decision model*

I am leading up to the point that an interpretable message, duly sent and received, will generally become information (i.e., stored in the mind of the receiver) if the receiver judges the message to be of some interest, relevance, or usefulness. This will typically occur if the message relates to a phenomenon about which the receiver wishes to make a judgement or decision and the result of that decision has potential utility.

Decisions typically have as their objectives the modification of the world in some fashion, or they represent alternative strategies to respond most effectively to different condi-

tions of the world. In order to do this most effectively, the decision maker should have some reasonably clear objectives and, furthermore, some utility function whose maximization represents a reasonable trade-off between the costs and benefits of alternative decisions. He/she is interested in those factors which have a relationship with or impact on this utility. The assessment of possible alternative decisions takes place within a formal or informal framework, or a thought process: the decision model. The decision model need not be a formal mathematical or probabilistic model, but it does involve an explicit or implicit utility function which is to be maximized or at least increased. The model can range from a complex econometric or simulation model to a loosely structured accumulation of experience. In fact, even where complex and formal models exist, few major decisions are made automatically on the basis of model predictions: typically these are tempered by judgements.

Consider three examples. The first involves a farmer's decision to apply X tons of fertilizer to a corn field to maximize net profits. The utility function can be well defined as the difference between the value of the expected increased yield and the cost of the fertilizer. Data on the cost and performance of fertilizers are directly relevant in this case, as would be data related to fluctuations in the price of corn. All these data qualify as potential information for the farmer – if he/she "receives" and "processes" these data. The second example involves the consideration of the alternatives to spend a fixed sum of money on a vacation or on increasing the value of one's house. The value to the person of the alternative decisions would be difficult to incorporate into a *formal* model. Yet people do take such decisions, so at least an implicit decision model is nevertheless involved. Part of this model may be more precisely formulated: for example, the trade-offs involved in the length, location, and level of luxury involved in alternative vacation plans. Clearly data about travel costs are relevant. Again, these various data and/or statistics represent potentially useful information for the traveller. A third example might relate to a governmental objective of reducing income inequalities. Here the utility function would be quite difficult to define, because the means to achieve such an objective are not fully understood and because important trade-offs have to be considered. Nevertheless, careful analysis of past experience, together with theoretical considerations, may identify several factors affecting income inequality: government transfer payments, tax policies, the general state of the economy, education, and family backgrounds. Statistics related to these would therefore be judged as relevant for the particular decision. They are transformed into information if interpreted within the context of the problem or within the analysis of alternative courses of action.

The world is infinitely complex – even the limited segment of it which a specific decision might be designed to modify or to which it is reacting. Necessarily, one can only cope with this complexity within the framework of a simplifying model. The (formal or informal) model helps in sorting out those factors which are expected to have some understood relationship with the outcome one wants to modify. Furthermore, the model would at least generally indicate how these relevant factors interrelate with one another and with the possible outcomes, which of the factors are capable of modification, what is the likely effect on the outcomes of modifying one or more of the factors, what is the current state of the relevant factors. Note that only the last of these questions can possibly be answered by statistical data; however, the past interrelationship among the factors can be explored through statistical analysis.

The decision maker, therefore, needs the

simplification offered by a formally articulated or informal model to cope with the complexities of the world. It also indicates which among the available data (interpretable messages) are relevant to the decision and should therefore be retained. From the point of view of the sender of an interpretable message, this role of the decisionmaker's model is particularly important. If the sender knows that his/her message relates to a factor which is part of the decisionmaker's model, he/she can be reasonably sure that the message will be interpreted. Thus, if the sender is able to highlight a relationship between his/her message and some factor of known interest to the receiver, then he/she can cause the data to be "received."

2.7. *Relevance*

As indicated above, data only has the potential of becoming information. If the utilization of data by a decision maker reduces the uncertainty associated with his/her decision, we say that the data is of relevance to him/her. Clearly, relevance is a property of the data in relation to a class of users or uses, not a property of the data alone.

A statistical agency wishing to render its data as widely relevant as possible must therefore acquire considerable knowledge of the decision issues and models of its users. Such knowledge is acquired through a variety of analytical activities shedding light on the users' decision problems – at the very least by maintaining close dialogues with a wide cross-section of users.

A significant determinant of relevance is the extent to which data fit the user's decision model. Appropriate disaggregation can often exponentially increase the explanatory power (relevance) of statistics. For example, data on the distribution of unemployment by age is often more relevent than the total number of unemployed. The ultimate disaggregation is provided by the underlying microdata base which indeed, is often found relevant by many users (assuming, of course, that the confidentiality of identifiable responses can be maintained). Furthermore, given the fact that most models have to use data from several sources, the usability of a given datum in a model strongly depends on the ease with which it can be used jointly with other data. Thus other prerequisites for increasing the relevance of data emerge: standardization of concepts, and ease of access of a variety of related statistics.

2.8. *Validity*

Data collection typically involves compromises between the concept a decision maker might wish to measure (the definitions he/she would have liked incorporated into the measurement process) and what is possible and practical to measure (the "operationalized concept"). Different users faced with different decision problems may well have different ideal concepts. However, those involved in conducting a survey must develop concepts which can be measured with a reasonable degree of reproducibility and which have a wide range of potential clients (each with potentially different "ideal" concepts). The concept actually measured is the "operationalized" concept. The distance between a given user's ideal concept and the operationalized concept measures the validity of the data for the given use. For example, the operationalized concept of unemployment used in the Canadian Labour Force Survey relates to labour market activity. Given a well developed program of unemployment insurance as well as the possibility of several job holders in the same family, it may not be suitable for monitoring the number of persons suffering economic hardship as a result of unemployment. Therefore the estimates have reduced validity for the latter purpose.

Relevance is a broader concept than validity – indeed the latter is subsumed in the for-

mer: relevance is reduced by the extent to which the user's ideal concept, as embodied in a decision model, is only approximated by the operationalized concept actually used. Validity, and hence relevance, can often be improved if the microdata base includes a sufficient range of variables so that different users can construct their own approximations of their ideal concept. For example, the Canadian Labour Force Survey includes a wide range of variables on labour market behaviour on the basis of which users can construct significantly different alternative definitions of unemployment (though all of them would relate to labour market behaviour, and none directly to resulting economic hardship).

2.9. Accuracy

The accuracy of data, broadly defined, is the extent to which the measurement of the operationalized concept is subject to error. It includes the well known components of measurement and, when applicable, sampling errors.

Accuracy is also affected by the extent to which the reference entity, in this case a group, is incorrectly identified – for example, by failing to include in the group persons who by definition belong to it. Accuracy which is not entirely adequate for a particular application reduces relevance. Put differently, accuracy commensurate with a given substantive objective is one of the many attributes of relevant data.

2.10. Potentially misleading data

The concept which has been measured is often described only very briefly (sometimes only through the use of a term like "unemployment.") In that case the receiver of the message may assume the concept to correspond to his or her ideal concept of unemployment. Similarly, unless an explicit statement about accuracy is provided, the receiver is free to impute any level of accuracy, including "complete accuracy." Thus data whose concepts and accuracy are inadequately or incompletely described are regarded as potentially misleading. Statistics Canada has an explicit policy on informing users about the errors in its statistical output as well as about the methodology used in its data collection and analysis.

3. Social Statistics in Canada

The concepts developed in the previous section will be used in an analysis of social statistics in Canada. It is a tentative case study of the impact of greater emphasis on analysis; a more concrete evaluation of the results of recent initiatives will have to wait for several years.

3.1. Predominance of administrative data in social statistics

Social statistics in Canada are dominated by data initially collected for administrative as opposed to statistical purposes. Vital statistics, statistics on health, educational, justice and correctional institutions evolved in response to specific decision problems, typically related to the administration of particular social programs. The programs themselves correspond to many of the major social policies of post-war Canadian governments. These programs were initiated using decision models that were more political than quantitative. Thus the process often relied as much on the articulation of programs as on clear objectives; and the analysis involved political and financial costs, as much as the alternative means to achieving well articulated objectives. Thus the launching or extension of programs, such as universal health insurance or the large-scale expansion of higher education, did not represent decision

problems[3] in support of which comprehensive statistical programs would have been developed.

Having launched the programs, their efficient administration entails continual decision-making that relies on statistics – administrative data and statistics from surveys. The definition of the concepts measured and the identification of the reference entities are determined to maximize their relevance given these particular decision models, e.g., the administration of hospital programs. But the problem of efficiently administering a set of hospitals is substantially different from the issue of how to improve the health of Canadians. Hospital statistics were largely developed in response to the administrative problem. This reduces their relevance for studies of the general health of Canadians, and the validity of the statistics might significantly be impaired for some of the latter analyses.

Neither the operationalized concepts, nor the reference entities (coverage) lend themselves easily to the development or assessment of general social policies – although a lot of work has been done by Statistics Canada to influence administrators to modify their concepts so that the validity is improved for more general uses. A further problem is the relative paucity of explanatory variables which could provide significant linkages to a variety of decision models and thereby enhance the relevance of the data that are available. The timeliness of the data for statistical purposes is also not uniformly good.

Attempts to overcome data limitations, for example, through record linkage or modifications to the administrative record keeping systems, are expensive and often controversial. Given that administrative data must be collected, additional statistical collections impose both extra costs and response burden. On the positive side, given that the cost of administrative data is ascribed to non-statistical purposes, the marginal cost of their statistical utilization is low. And, often enforcement or incentive mechanisms are available to improve their accuracy, i.e., their accuracy as measures of operationalized concepts.

Since the question of extended relevance of data is, of necessity, of secondary importance to administrative agencies, these agencies typically give inadequate recognition to the usability of the resulting statistics in decision models other than those they were initially designed to support. Nevertheless, given the relevance of administrative data sources for a broad class of users, overcoming the data's shortcomings for non-administrative uses is a major challenge for Statistics Canada, one to which we attach a high priority (Brackstone (1987)).

3.2. *Social Surveys*

Much statistical information is not available as a by-product of the administrative processes and can be obtained only through surveys. This is particularly true for analyses involving decision models which cut across the subject boundaries imposed by the institutional boundaries of administrative processes.

As pointed out earlier, the relevance of data is partly determined by their accuracy. The requirements of accuracy, particularly if subnational analyses are required, often impose a sample size which renders data collection prohibitively expensive for the private sector. At the Federal level, the main vehicles used to collect social statistics are: the quinquennial censuses of population and housing (Pryor (1986), Statistics Canada (1987), the Labour Force Survey (Statistics Canada (f), monthly), the annual Consumer Finance and Household Facilities Surveys (Statistics

[3] In recent years there has been more interest and activity in the area of program evaluation, particularly before changes to major programs are introduced.

Canada (c), annual), the biennial Family Expenditures Survey (Statistics Canada (a), occasional), and supplements attached to the Labour Force Survey (Statistics Canada (e), annual). During 1978–79 a new social survey, the Canada Health Survey (Health and Welfare Canada and Statistics Canada (1981)), was launched but was a victim of budget cuts. A major post-censal Health and Activity Limitation Survey (Furrie (1987)) was carried out following our 1986 census directed at those who were identified as "handicapped" by some screening questions included in the census. Recently we launched a small-scale annual General Social Survey (Statistics Canada (1986)) to which I shall return in a subsequent section.

Canada has a quinquennial census of population and housing. Given our relatively underdeveloped program of social surveys, the great importance of subnational breakdowns in large countries, and the great interest in relatively small population groups (e.g., the poor, the elderly, ethnic groups, the disabled, etc.), censuses shoulder a heavy burden within our social and socio-economic statistics. The priority attached to censuses by our clients has twice been confirmed in recent years, once when the government considered (but subsequently abandoned) the idea of restricting the 1981 census to its legally mandated minimum content, and again when it cancelled (and subsequently reinstated) the 1986 census. On both occasions the change was brought about by widespread and effective user reaction.

One can ask whether our current survey program is big, small or just about right. It is certainly small in comparison with U.S. social statistics. We do not have regular surveys on health, comprehensive victimization, housing, time allocation, quality of life, longitudinal surveys to measure the impact of income changes or to monitor special cohorts which are in a state of significant transition (e.g., the pre-retirement group, or entrants into the labour force). This partial list of data gaps could be extended considerably. However, such abstract questioning of the adequacy of our survey program is probably unproductive. Germane questions would be: what are the decision models which cannot be formulated and what important analyses cannot be carried out with an acceptable level of uncertainty without such data? It is through relevant analytical work that social scientists, very much including those in the statistical offices, can help in answering these questions.

3.3. *Analysis of social statistics*

In an outstanding paper delivered at the Joint Statistical Meetings of the American Statistical Association, Robert Parke and Eleanor Sheldon (1973) identified several types of analyses of social statistics serving to render the statistics relevant for major policy purposes. I would like to recapitulate briefly the analytical categories they identified, using my own terminology and Canadian examples, to illustrate the wide variety of important social decision models which can be assisted by suitably analyzed social data or, in fact, which can be modified by such data and analyses.

3.3.1. Cognitive information

Studies which identify important new social problems and map new social terrain with fresh information, fall into this category. Illustrative examples are the studies of: the lack of overall change in Canada's income distribution over two decades despite *major redistribution measures* (Wolfson (1986 a, b); and Statistics Canada (d)); the central role of acquaintanceship in violent offences (Statistics Canada (b), annual); the interaction of legal process and personal characteristics in deter-

mining the outcome of divorce proceedings (McKie, Prentice, and Reed (1983)); the degree of disability inherent in different handicaps (Statistics Canada and Department of the Secretary of State (1986)). All of these studies brought about new understandings of intrinsic interest, and also of direct use to those responsible for social policies. Such studies may indirectly result in the formulation of more specific, problem-oriented decision models.

3.3.2. Identifying important external constraints

The contribution here is to invite the attention of policy makers to developments not under their control that may alter the way they conduct affairs that are under their control. An example from outside the social sciences is the weather reports in which we are interested because they may alter our "adaptational strategy," even though we cannot manipulate the weather. Within the social sciences, there are numerous studies showing the impact of demographic trends on: the female labour force (Statistics Canada and Employment and Immigration Canada (1981)), income distribution (Wolfson (1986 a, b)), linguistic and other minority groups (Harrison and Lachapelle (1985)), geographic mobility of workers (Shaw (1985)), unemployment, GNP, consumption patterns, schools, health services (Angus, Lefebvre, and Strohmenger (1982)). Specific policies can affect each of these separate problem areas, but only within a margin determined by external constraints. The setting of realistic goals and the definition of success or failure of particular policies is an important indirect contribution.

3.3.3. Projecting consequences

This category is somewhat similar to the previous one with the exception that manipulative, as opposed to adaptive, strategies are available. An illustrative example is a study of the combined factors of the changing age structure of Canada's population (the "graying" of Canada), changing morbidity patterns, and utilization of health services (Stone and Fletcher (1982)). Statistics on these phenomena have been analyzed to project how present trends would translate into demand for the several main types of hospital care over the next few decades. The analytical results clearly show the need for adaptive actions to prevent a serious dislocation in the provision of hospital services.

3.3.4. Analysis of specific decision options

These analyses evaluate how well different decision options achieve their goals. One example is the simulation model developed by Statistics Canada (Wolfson (1986 a, b)) and used to evaluate the impact of alternative tax and income transfer policies on the financial situation of different types of households, and to assess the range of possible consequences of various unemployment insurance policies on labour force participation.

3.3.5. Communicating the meaning of data

Through data analysis, statisticians try to illuminate trends and relationships inherent in the data. This type of analysis is a minimal requirement to transform the data into information; without it the data are in danger of remaining an interpretable message. This analysis indirectly highlights the potential relevance of data for a variety of not necessarily explicitly specified decision models. Examples are: highlighting trends and apparent changes in trends (Reed 1984)); the transformation of mortality data into life expectancy tables (Nagnur (1986)); the presentation of income data in the form of measures of income inequalities, trends in real

incomes, trends in the relative importance of different sources of income (Wolfson (1987)); the transformation of current marriage and divorce data into cohort data reflecting different divorce experiences (Adams and Nagnur (1981)); standardization of populations over time with respect to some characteristics to study the impact of other characteristics (Wolfson (1986)); highlighting trends of selected ratios, e.g., the ratio of manufacturers' current inventories to new orders received. Some important analyses can be accomplished using very simple tools, for example, "making sense" of graphs.

In spite of these illustrations, the amount of Canadian social statistics analysis falls far short of its potential. Some needed analyses are perhaps not available because the required data are unavailable. Unfortunately, however, there appears to be a shortage of analyses even when the statistics are available. Most of our household survey data are, for example, available on tape in micro-data form – ready for analyses. The Labour Force Survey or the income surveys are examples. Yet the published literature contains relatively few examples of their utilization.

3.4. A climate of scarce resources

A climate of extremely scarce, indeed diminishing, resources predominates in the Canadian government service and certainly within Statistics Canada.

Government priorities can be influenced either by demonstrating how the government's own effectiveness depends on a certain level and type of statistical service, and by illustrating how this service is essential for other levels of government or groups of the population which, in turn, are considered to be important to the government. In either case, *demonstrations* provided by case studies of the use of statistics for important purposes are needed. But, with the possible exception of legislated or other "formula type" use of statistics, most other important uses involve analysis and decision models – broadly interpreted in the sense of Section 2. Indeed, even the formula type uses were typically preceded by analysis.

Thus, while undoubtedly some needed analyses are not carried out because the relevant data are missing, an equally convincing case can be made that a wide range of needed social (and other) statistics are not funded because of the relatively low volume of analysis in this area.

I believe social science itself must become more relevant to decision makers in order for social statistics to be given higher priority. More analyses are required to identify social problems of recognized importance; to define factors thought to be related to such problems, including those which can be influenced through decisions; to articulate the appropriate decision models and to determine the statistical information which is missing; and finally to communicate effectively the results of the analysis not only to colleagues but also to decision makers and the general public.

4. Some Recent Initiatives

Statistics Canada believes that it is in the national interest to encourage more analysis of existing statistics to increase their relevance, and identify gaps which hinder a better understanding of important public issues. A number of concrete initiatives have been taken in recent years to strengthen analytic activities by agency staff, while at the same time maintaining our tradition of scientific objectivity; and to encourage analytic activities outside the agency.

4.1. Analytic studies branch

A new Analytic Studies Branch has been established. Members of the branch are free

from routine operational commitments. The initial "critical mass" was obtained by bringing together some existing analytic units, supplemented by a small number of judiciously selected senior external appointments.

The branch has no intention to suggest a centralization of analytic studies. Rather, through a variety of collaborative arrangements with other divisions, the branch is to provide bureau-wide leadership and impetus. Indeed, by establishing a network of external relationships, the branch is to play an important role in strengthening analytic activities outside Statistics Canada.

The head of the branch is a senior officer who can, and is expected to, act as a spokesperson within Statistics Canada for the needs of the internal and external analyst community: with respect to data developments, internal computing facilities, and dissemination strategies.

The publicly available results of the work of the branch include papers, monographs, as well as formal analytic models.

4.2. *Professional advisory committees*

Statistics Canada established some 20 professional advisory committees to advise on specific subject areas (e.g., prices, labour, demography, external trade), and functions (e.g., statistical methodology, analysis). The committees are non-representational, membership is based on individual professional standing. Several of the committees have taken an interest in the analytic program of a specific area and help improve the cooperation between Statistics Canada and academic and other non-governmental research staffs. In addition, the Advisory Committee on Analysis and Research provides strategic guidance to the analytic program as a whole.

4.3. *The General Social Survey*

By reducing the scale of some of our other current survey operations, budgetary provision was made for an annual General Social Survey (Statistics Canada (1986)). Each year the survey will be devoted to an important social policy issue. As such, the selection of each topic area will explicitly be driven by analytic considerations.

Past topics included: family dynamics as reflected in retrospective interviews with females regarding the spacing of marriage, divorce, children, education, labour market participation; formal and informal support systems sustaining the elderly population; language acquisition, use, and transfer; and health status.

We are seeking to engage academics under contract, or as part of their sabbatical work, to collaborate with us in articulating the appropriate conceptual frameworks for questionnaire development, and to carry out substantive analyses of the data that has been collected. Tha data will also be released on computer tapes in the form of microdata, to encourage secondary analyses.

4.4. *Fellowship program*

A Fellowship Program has been established to attract visiting scholars to Statistics Canada. Fellows are encouraged to come to Statistics Canada for a period ranging from several months to a year and exploit our available statistics and data processing support in the pursuit of jointly identified analytic or methodological objectives. It is hoped that the fellows will both stimulate our in-house staff and, in turn, will leave the bureau much better informed regarding the analytical potential of available data, including both strengths and weaknesses.

4.5. Program evaluation

A formal program evaluation was initiated. This involves a review, on a rotational basis, of the major programs of the agency. This review concentrates on client perspectives: it probes the continuing relevance of programs, ways to increase program usefulness (e.g., improved dissemination), and the elimination of statistical gaps which should have highest priority.

The evaluations, as a matter of policy, are carried out by external experts under contract, managed by an in-house independent program evaluation group. It is intended that the assessments should be analytic, i.e., not only list clients who express a need for some statistical program but also articulate the substantive nature of their requirements.

It is our experience that such evaluations generate greater awareness of the substantive issues for which our clients require statistical information. This, in turn, is a prerequisite for maintaining the relevance of statistical programs.

4.6. Release of statistics

Statistics Canada has, for many years, published a daily bulletin called *The Daily* which serves as the vehicle of first release of all statistics: no statistical data can be released before its availability is announced in *The Daily*. This policy is designed to ensure equal access to official statistics by all users. It also serves as a marketing mechanism since releases of *The Daily* is quite eagerly followed by the media as a convenient summary of new statistics.

Recently, a determined effort has been made to improve the text of the releases published in *The Daily*. I am conducting a personal campaign to ensure that all releases in *The Daily* are both "readable" and analytical. This includes emphasis on significant trends, comparisons, graphs, and a virtual embargo on mechanical verbalization of published tables. This will undoubtedly improve the value of *The Daily* as a means of placing summary statistical information into the public domain via the media. At the same time, we hope that the routine requirement of basic analysis will have a broad cumulative effect on the skills of the professional staff.

A weekly summary version of *The Daily*, called *Infomat*, is designed with busy but well informed generalists in mind (business leaders, senior government officials, and academics with broad interests). Its front page includes a set of headlines and a one sentence summary of all major statistical releases of the week. Corresponding to each headline is a brief article consisting of a few paragraphs, a chart or graph, and perhaps a single highly summarized table. Each article ends with the appropriate reference (and telephone number for contact) for those wishing more detailed data.

An initiative closely related to that descibed above is a new policy requiring that all statistical publications contain a highlights section. The highlights are designed to enhance the relevance of publications and to force subject matter personnel to carry out routinely a basic level of data analysis.

The strong emphasis on increased analytic activities has to be balanced by appropriate safeguards to ensure that the analytic output is both professionally sound and of a character appropriate for the statistical agency. With this in mind, we promulgated a policy on institutional and peer review applicable to all analytic output. Briefly, the policy requires a two-pronged review: institutional review carried out by senior managers to ensure that the subject of the analysis is relevant and the commentary preserves institutional objectivity (avoid policy advocacy, state all assumptions underlying model-derived results, etc.); and peer review, carried out by professional experts from outside and

inside the agency to assess the scientific validity of the work.

5. Concluding Remarks

Analytic activities are not new for Statistics Canada or, indeed, for most statistical agencies. For example, the compilation of the System of National Accounts involves a great deal of analysis. At a more elementary level, the selection of what should be published and the design of statistical tables necessarily involves analytic assessments.

What is, perhaps, relatively new is the increased emphasis on data analysis (i.e., exploration of a given body of data), issue oriented analysis (i.e., exploration of how statistical analysis can shed light on a given societal issue), and effective communication of the results of the analyses.

At a time of significant budget constraints (indeed of declining resources) it might superficially appear unwise to maintain or even marginally increase the emphasis on analysis. Yet, it seems that to do so is perhaps even more important now than in times of increasing resources. To the extent that analysis enhances the relevance of our output, it more than repays a modest investment, particularly because data collection is so much more expensive.

It is not easy to publicly defend such an approach, partly because analysis appears expendable and marginal to our mandate, partly because (in contrast to large-scale national data collection activities) it can and is carried out by researchers from outside the statistical office as well. But I believe that analysis is not expendable, that some amount of analytic activity by Statistics Canada staff is a significant facilitator for outside researchers, and that its feedback to programs, priorities, and staff development are essential for the long term well-being of the agency.

Analytic activities require significant management attention to facilitate their proper development, to ensure the necessary interactions both with other internal staff and with outside peers, and to ensure that the activity remains within bounds which are commensurate with the objective non-political image of the statistical office.

6. References

Adams, O.B. and Nagnur, D.N. (1981): Marriage, Divorce and Mortality: A Life Table Analysis for Canada, 1975–77. Catalogue 84–536, Statistics Canada, Ottawa.

Angus, D.E., Lefebvre, L.A., and Strohmenger, C. (1982): An Analysis of Hospital Expenditures in Canada. Catalogue 83–522E, Statistics Canada, Ottawa.

Brackstone, G.J. (1987): Issues in the Use of Administrative Records for Statistical Purposes. Survey Methodology, 13, (1), pp. 29–43.

Furrie, A. (1987): The Health and Activity Limitation Survey. A National Database on Disabled Persons — Making Disability Data Available to Users. Uncatalogued, Statistics Canada, Ottawa.

Health and Welfare Canada and Statistics Canada (1981): The Health of Canadians: Report of the Canada Health Survey. Catalogue 82–538E, Ottawa.

Harrison, B.R. and Lachapelle, R. (1985): Language Maintenance and Language Shift in Canada. Migration Today, Vol. XIII, 4, pp. 13–17.

McKie, D.C., Prentice, B., and Reed, P. (1983): Divorce: Law and the Family in Canada. Catalogue 89–502E, Statistics Canada, Ottawa.

Nagnur, D. (1986): Longevity and Historical Life Tables, 1921–1981 (Abridged), Canada and the Provinces. Catalogue 89–506, Statistics Canada, Ottawa.

Parke, R. and Sheldon, E.B. (1973): Social Statistics for Public Policy. American Sta-

tistical Association, Proceedings of the Social Statistics Section, pp. 105–112.

Pryor, E.T. (1986): Approaching the 1991 Census of Canada. Second Annual Research Conference, U.S. Bureau of the Census, Reston, Virginia, March 23–26.

Reed, P. (1984): Historical Statistics on Justice. Uncatalogued, Statistics Canada, Ottawa.

Shaw, R.P. (1985): Intermetropolitan Migration in Canada: Changing Determinants Over Three Decades. N.C. Press Ltd. and Statistics Canada, Toronto.

Statistics Canada (1987): Proceedings of the International 1991 Census Planning Conference. Ottawa, October 8–11, 1985.

Statistics Canada (1986): General Social Survey – Status Report. Uncatalogued, Ottawa.

Statistics Canada (a): Family Expenditure in Canada. Catalogue 62–555, occasional, Ottawa.

Statistics Canada (b): Homicide in Canada: A Statistical Perspective. Catalogue 85–209, annual, Ottawa.

Statistics Canada (c): Household Facilities and Equipment. Catalogue 64–202, annual, Ottawa.

Statistics Canada (d): Income Distributions by Size in Canada. Catalogue 13–207, annual, Ottawa.

Statistics Canada (e): Overview of Special Surveys. Uncatalogued, annual, Ottawa.

Statistics Canada (f): The Labour Force. Catalogue 71–001, monthly, Ottawa.

Statistics Canada and Department of the Secretary of State (1986): Report of the Canadian Health and Disability Survey 1983–84. Catalogue 82–555E, Ottawa.

Statistics Canada and Employment and Immigration Canada (1981): Demographic Trends and Their Impact on the Canadian Labour Market. Catalogue 71–603E, Ottawa.

Stone, L.O. and Fletcher, S. (1982): A Profile of Canada's Older Population. The Institute for Research and Public Policy, Montreal.

Wolfson, M. (1986a): Stasis Amid Change – Income Inequality in Canada 1965–1983. Review of Income and Wealth, December, pp. 337–369.

Wolfson, M. (1986b): A Guaranteed Income. Policy Options, January, pp. 35–45.

Wolfson, M. (1987): Homemaker Pensions and Lifetime Redistribution. International Association for Research in Income and Wealth, Twentieth General Conference, Rocca di Papa, Rome, Italy, August 23–29.

Received September 1986
Revised January 1988